Real-time Estimation of Physical Activity Intensity for Daily Living

L. Tao, T. Burghardt, M. Mirmehdi, D. Damen, A. Cooper, M. Camplani, S. Hannuna, A. Paiement, I. Craddock

SPHERE, Faculty of Engineering, University of Bristol, Bristol, UK, BS8 1UB

Keywords: Assistive monitoring, computer vision, energy expenditure, activities of daily living

Abstract

Estimating a person's energy expenditure and activity intensity over time is an important component in managing various health conditions or tracking lifestyle choices. To implement an automatic estimation, most current systems ultimately require users to wear sensor devices. In contrast, this paper presents a framework for the contact-free, real-time estimation of energy expenditure, applicable to daily living scenarios. This is a new application in real-time computer vision. We demonstrate the effectiveness and the benefits of utilising a basic set of features and evaluate the resulting framework on the challenging *SPHERE-calorie* dataset. To ensure accurate evaluation, automated estimates are compared against a simultaneously taken indirect calorimetry ground truth based on per breath gas exchange. Following detailed experiments, we conclude that the proposed real-time vision pipeline is suitable for monitoring physical activity levels in a controlled environment with higher accuracy than the commonly used manual estimation via metabolic lookup tables (METs), whilst being significantly faster than existing automated methods.

1 Introduction

Physical activity is an important determinant in understanding the development of chronic diseases. Current evidence-based guidelines [19] indicate that people who are regularly physically active have a 20% to 40% lower risk of developing conditions such as cardiovascular disease and type 2 diabetes than those who are inactive, and suggest that adults should accumulate at least 150 minutes of moderate intensity physical activity each week or 75 minutes of vigorous activity, or a combination of the two.

Energy expenditure, also referred to as 'calorific expenditure', is one commonly used single metric to quantify physical activity levels over time. It can be accurately measured using a calorimeter which operates based on the respiratory differences of oxygen and carbon dioxide in the inhaled and exhaled air. Measurements can either be direct via a sealed respiratory chamber [17] or indirect which requires carrying gas sensors and wearing a breathing mask [1]. However, these devices are impractical to use routinely in daily life due to their high cost, lack of portability and cumbersomeness. On the other hand, wearable devices have become a popular choice to measure coarse categorisation of activity intensity levels [5]. Among these, tri-axial accelerometers are the most broadly used inertial sensors [8].

Computer vision techniques that help with the diagnosis and management of health and wellbeing conditions have also started to draw some research attention recently. Yet, although there exists a significant body of literature describing the inference of activities from 2D colour intensity imagery [2], RGB-D data [3], and skeleton-based data [16], studies on energy expenditure using visual sensors have been relatively limited. RGB only video has recently been used by Edgcomb and Vahid [10] to coarsely estimate daily energy expenditure where a subject was segmented from the scene background and changes in height and width of the subject's motion bounding box, together with vertical and horizontal velocities and accelerations, were then used to estimate calorific uptake. However, we note that their regression models were trained based on the ground truth readings reported by wearable accelerometry, which may provide only an approximate benchmark. Tsou and Wu [24] took this idea further and estimated calorie consumption using full 3D joint movements tracked as skeleton models using a Microsoft Kinect. In this setting skeleton data is commonly noisy and currently only operates reliably when the subject faces the camera [23], thus the method has difficulties to generalise to more unconstrained scenarios.

The above examples exemplify that calorific uptake and linked activity levels can often be directly related to body motion. Motion information could also be recovered directly using the optical flow derived from two adjacent RGB images [22] or 4D surface normals [14] and more recently, dense 3D flow [20] from depth images. These approaches, however, often suffer from unaffordable run time, for example with a reported computation time of up to 9 minutes per frames in [20].

Apart from the currently performed activity, energy values are also highly dependent on the previous energy expenditure, as adaptations of the human body cause an exponential increase/decrease to a plateau in oxygen consumption until a steady state corresponding to the current activity is attained [12]. Therefore, the motion information needs to be recovered from a sequence of data over some time window to infer calorific uptake levels. Concatenating per-frame descriptors is straightforward, but it often suffers from the curse of dimensionality and related high computational cost. Compacting data within a temporal window may be achieved to some degree by abstracting large feature arrays [15, 18], but remains a challenge. Thus, in essence, any system will require capturing

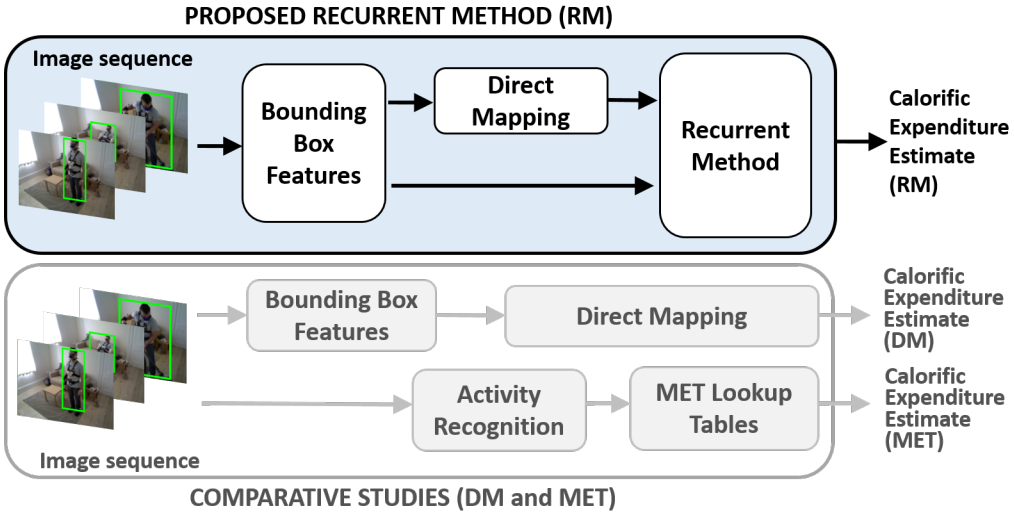

Figure 1. **Framework Overview.** Image sequences are represented by features extracted from bounding boxes. The proposed recurrent method RM (top) then maps features to calorie estimates. We compare this method to a direct mapping method DM, and a manual standard using a lookup table MET (bottom).

visual aspects relevant to calorific expenditure, whilst limiting the dimensionality of the descriptor.

Here, we propose a real-time framework for estimating calorific expenditure levels from bounding box features only. We evaluate the proposed system over daily activities performed in a living room environment. Figure 1 shows in bold a flowchart of our proposed approach – extracting features from the bounding box, mapping the features directly to calorie estimates via a monolithic classifier, and adding a cascaded and recurrent classifier as the last step to capture temporal dependencies (RM in short). The proposed method is compared against a ground truth as-exchange measurements (GT in short) and two alternative methods also shown Figure 1: (1) direct mapping to calorie estimates without recurrent approach (DM in short), and (2) manual mapping from activity classes to calorie estimates via the Metabolic Equivalent Task lookup tables [4] (MET in short). We also compare the processing time of feature extraction and estimation accuracy of the proposed method against that of a fully fledged vision system [21], which uses detailed flow and depth features at the cost of sacrificing real-time capabilities. We will show that the proposed system can operate under real-time constraints whilst achieving accurate activity intensity level prediction outperforming the widely used MET method.

2 Proposed Method

2.1 Feature Representation

We first extract base features from bounding boxes, and then form higher level motion features by a set of temporal filters. We use the bounding box returned by the OpenNI SDK [13] person detector and tracker using an Asus Xtion for capture. Our per-frame descriptor describes the velocity vector and the ratio of height and width of the bounding box.

To represent both short and long term temporal changes in a video, one may model how the local/global information is changing over time. Pooled motion features were first presented in [18], designed for egocentric video analysis. We modify this pooling operator to make it more suitable for our data.

Figure 2 illustrates the overall process of feature assembly covering the initial time series representation, followed by temporal pyramid alignment, and the final, serialised representation of the descriptor vector.

Let $\mathbf{S}$ be a set of time series data, such that $\mathbf{S} = \{S_1, \ldots, S_N\}$, $\mathbf{S} \in \mathbb{R}^{N \times T}$ for a video in matrix form, where N is the length of the per-frame feature vector, and T is the number of frames. A time series $S_n = [s_n(1), \ldots, s_n(T)]$ is the n^{th} feature across $1, \ldots, T$ frames, where $s_n(t)$ denotes n^{th} feature at frame t. The time series data $\mathbf{S}$ consists of a set of time segments as $\mathbf{S} = [\mathbf{S}_i^1, \ldots, \mathbf{S}_i^{2^i}]$ at level i. A set of temporal filters with multiple pooling operators is applied to each time segment $[t_{min}, t_{max}]$ and produces a single feature vector for each segment via concatenation.

As Figure 2 illustrates, we use three pooling operators, that is max pooling, sum pooling, and spectral pooling. The first two are defined respectively as

$$\mathcal{O}_{\max}(S_n) = \max_{t=t_{\min} \cdots t_{\max}} s_n(t) \tag{1}$$

$$\mathcal{O}_{\text{sum}}(S_n) = \sum_{t=t_{\min}}^{t_{max}} s_n(t). \tag{2}$$

Spectral pooling is used to perform dimensionality reduction of the time series S_n in the frequency domain by the discrete cosine transform and then truncating the representation. The pooling operator takes the absolute value of the j lowest frequency components of the frequency coefficients D, in order

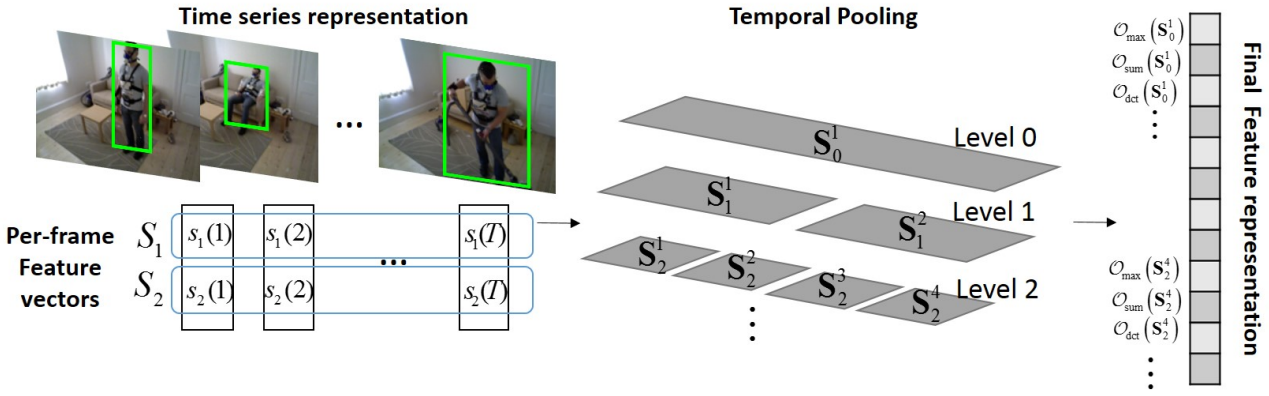

Figure 2. **Per-frame Feature Representation, Temporal Pyramid Pooling and its Feature Representation.** This schematic shows the temporal subdivision of data into various pyramidal levels (middle) and the concatenation of resulting features (e.g. max, sum and DCT) into a descriptor vector (right).

to help remove high frequency noise

$$\mathcal{O}_{\text{dct}}(S_n) = |M_{1:j} S_n|\,, \tag{3}$$

where M is the discrete cosine transformation matrix.

2.2 Recurrency

We pose the energy expenditure estimation problem as a sequential and supervised regression task. We train a support vector regressor to predict calorie values from the given features over a training set. The sliding window method naturally converts the sequential supervised learning problem into the classical supervised learning problem, which maps each input window of width w to an individual output value y_t. The window contains the current and the previous $w-1$ observations. The window features are assembled by temporal pooling from the time series $\mathbf{S} = \{S_{t-w+1}, \ldots, S_t\}$.

The energy values for a particular time are highly dependent on the energy expenditure history, thus the sliding window methods can be extended by including recurrent information. In our system, these are most directly expressed by previous calorific predictions during operation. Thus, employing recurrent sliding windows offers an option to not only use the features within a window, but also take the most recent d predictions $\{\hat{y}_{t-d}, \ldots, \hat{y}_{t-1}\}$ into consideration to help predict y_t. During learning, as suggested in [9], the ground truth labels in the training set are used in place of recurrent values.

3 Experimental Results

3.1 Dataset and Parameter Settings

In order to quantify performance of the proposed approach, we conducted experiments on the *SPHERE-calorie* dataset[1] [21]. It is a very challenging dataset for calorific expenditure estimation collected within a home environment covering daily living activities. The dataset consists of an RGB-D video sequences captured by a Asus Xtion camera mounted at the corner of a living room and ground truth readings from a COSMED K4b2 [1] portable metabolic measurement system. The dataset was generated over 20 sessions by 10 subjects with varying anthropometric measurements containing up to 11 common household activity categories per session. Each session lasts around 30 minutes, and totalling around 10 hours recording time. To reflect variations in transitions between activity levels, we consider 9 different combinations of three activity intensities in each session.

Colour and depth images were acquired at a rate of 30Hz. The calorimeter gives readings per breath, which occur approximately every 3 seconds. Figure 3 shows a detailed example of calorimeter readings and associated sample RGB images from the dataset, together with activity intensity levels. The raw breath data is noisy (in red), and so we apply an average filter with a span of approximately 20 breaths (in blue). The participants were asked to perform the scripted activities based on their own living habits without any extra instructions.

The categories and their associated MET values (in brackets) are: *sit still (1.3), stand still (1.3), lying down (1.3), reading (1.5), walking (2.0), wiping table (2.3), cleaning floor stain (3.0), vacuuming (3.3), sweeping floor (3.3), squatting (5.0), upper body exercise (4.0).*

We compare the proposed method RM to the direct mapping method DM and the Metabolic Equivalent Task method MET. DM is formalised as $Y_t = f(X_t)$, where Y_t is the target calorie value regardless of activity at time t, and X_t contains the associated feature vector over a window. The goal is to find a function $f(\mathbf{.})$ that best predicts Y_t from training data X_t. MET, one of the widely used methods for recording of the intensity of a physical activity by clinicians and physiotherapists, assumes that the clusters of activity are known. A MET value is assigned to each cluster, together with anthropometric characteristics of individuals. The amount of activity-specific energy expended can then be estimated as $energy = 0.0175(kcal/kg/min) \times weight(kg) \times$ MET values [4].

According to [4], activity can be categorized into three different intensity levels based on either MET values for each activity or the average energy consumed per minute. Table 1 outlines the activity intensity levels and their associated energy expenditure ranges. In our experiments, activity intensity levels

[1]The dataset is released on SPHERE website `http://www.irc-sphere.ac.uk/work-package-2/calorie`

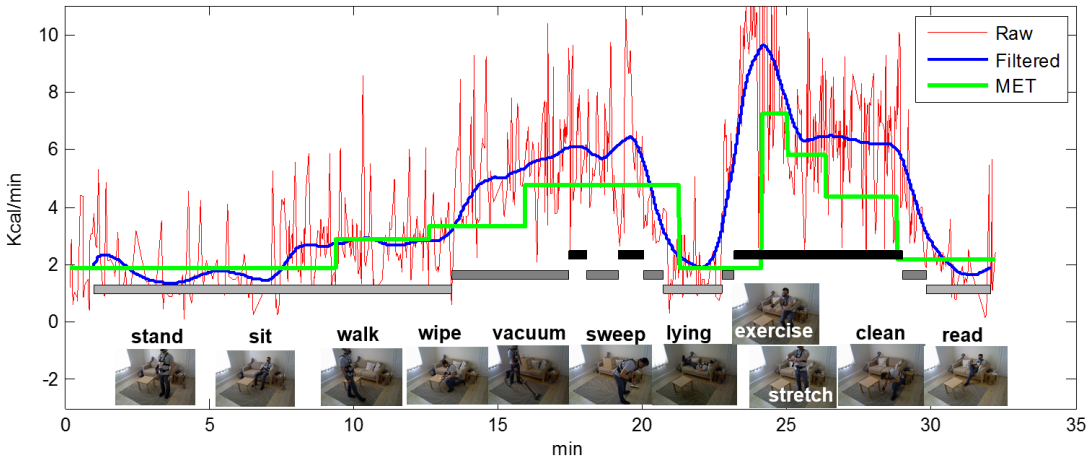

Figure 3. **Ground Truth Example Sequence.** Raw per breath data (red), smoothed COSMED-K4b2 calorimeter readings (blue), predicted calorie values using MET table (green), and sample colour images corresponding to the activities performed by the subject. Activity intensity levels are quantised into three levels based on ground truth readings (gray bars).

are quantised based on the ground truth readings (see Figure 3) instead of using MET values. This is because a fixed number is assigned to each activity which may overlook the drift during activity and transition between activities.

We use a linear support vector regressor for predicting calorie values from training data. The libsvm [7] implementation was used in the experiments. For testing, we apply leave-one-subject-out cross validation on the dataset.

This process iterates through all subjects, and the average testing errors of all iterations are reported. We use the root-mean-squared error (RMSE) as a standard evaluation metric for the deviation of estimated calories from the ground truth.

	MET values	kcal/min
Light	< 3.0	< 3.5
Moderate	3.0 - 6.0	3.5 - 7
Vigorous	> 6.0	> 7

Table 1. **Physical Activity Intensity Levels.** The table shows intensity levels and their associated energy expenditure ranges.

3.2 Quantitative Evaluation

Temporal Window Size - The accuracy of predicted calorie values is affected by the number of previous frames used for making the prediction. For the first set of experiments, we use the direct mapping method DM to investigate the relation between window length and calorie prediction errors.

All the sequences are tested with three different window sizes $w = \{450, 900, 1800\}$, corresponding to a 15, 30 and 60 seconds time slot. Table 2 illustrates the average RMSEs for calorie prediction of different window length w.

The results clearly show that calorie values are better predicted when the larger window (60 seconds) is applied. This may be attributed to the fact that human body adaptation causes an adjustment [12] of energy uptake over significant time durations, and thus access to previous measurements becomes a vital cue for accurately predicting current consumption.

Evaluation of recurrent system layout - We set the direct mapping method DM with window size $w = 1800$ as our baseline method. To evaluate the use of recurrency, we now test two recurrent sliding window approaches to explicitly encode previous energy estimates. The first one (RM1) uses the most recent predictions of the baseline method as input together with both visual features to predict current calorie value. Thus, it implements indirect recurrency utilising the predicted values from the baseline as recent predictions. The second one (RM2) implements full recurrency, i.e. it uses its own output as recurrent input together with visual features.

Table 3 shows the effect of using recurrent information. The best results for each activity are highlighted. In general, RM1 outperforms the other approaches for most activities. As expected, a recurrent method captures information that was not only being captured by the current sliding window. However, the full recurrency, RM2, suffers from significant drift and produces the worst results for half of the activities and also overall. We select RM1 as the "the proposed method RM" in the following sections.

Model Comparison - Table 6 provides the results for each sequence of the dataset individually. We report estimation accuracy and also the correlation between the ground truth and the observed values. The proposed RM achieves higher accuracy and correlation in more sequences than DM and MET based methods, and obtains better rates on average. In addition, we compare performance to a system using complex visual features (VF in short) [21] instead of bounding box features only. VF produces, as expected, better prediction results in most cases, however, RM operates more than 400 times faster than VF as detailed in the next section.

Looking at coarse categories of calorific expenditure, Table 4 lists overall results for the accuracy of predicting activity intensity levels only. For this task it is worth noticing that the proposed RM is able to produce comparable results to VF.

Processing Time - To analyse the efficiency of the proposed method further, we compare the processing time of feature construction procedures in RM and in VF where the average runtime of each frame though all subjects is reported. All the re-

w	*stand*	*sit*	*walk*	*wipe*	*vacuum*	*sweep*	*lying*	*exercise*	*stretch*	*clean*	*read*	**overall**
450	0.68	0.76	0.92	0.84	1.44	1.79	1.26	2.78	2.96	1.53	1.17	1.40
900	0.65	0.72	0.92	0.83	1.45	1.77	1.22	2.55	2.86	**1.53**	1.15	1.36
1800	**0.60**	**0.68**	**0.90**	**0.80**	**1.40**	**1.77**	**1.20**	**2.33**	**2.56**	1.56	**1.07**	**1.33**

Table 2. **Temporal Window Size and Calorific Expenditure Prediction.** Calorific expenditure prediction error (RMSE) with different window length. The best results in each activity are in bold.

	stand	*sit*	*walk*	*wipe*	*vacuum*	*sweep*	*lying*	*exercise*	*stretch*	*clean*	*read*	**overall**
DM	0.60	0.68	0.90	0.80	1.40	1.77	1.20	2.33	2.56	1.56	1.07	1.33
RM1	**0.57**	**0.62**	0.87	**0.80**	1.39	**1.66**	**1.11**	**2.30**	**2.21**	1.48	**0.95**	**1.24**
RM2	0.58	0.62	**0.82**	0.96	**1.31**	1.81	1.33	3.67	2.57	**1.17**	1.17	1.40

Table 3. **Activities and Calorific Expenditure Prediction.** Average calorific expenditure prediction errors (RMSE) for each activity with different learning approaches. The best results in each activity are in bold.

	RM	DM	MET	VF
Light	86.65	90.68	89.60	*85.02*
Moderate	83.81	72.58	54.21	*86.99*
Vigorous	79.91	59.32	40.24	*86.32*
Overall	84.79	81.85	75.94	*85.38*

Table 4. **Estimation of Activity Intensity Levels.** Recognition accuracy (%) of activity intensity levels. The table shows the performance of the proposed real-time method RM, compared to other approaches.

	feature extraction (ms)	temporal pooling (ms)	overall (ms)
RM	7.213E-04	1.4	1.4
VF	446.6	3.3	449.9

Table 5. **Runtime Performance Results.** Average computational costs (in milliseconds) for each frame processed by the VF and DM methods.

sults are produced using Matlab on a workstation with an Intel i7-3770S CPU 3.1GHz processor and 8Gb RAM. Table 5 shows the average computational costs for feature extraction, temporal pooling and overall costs of each frame.

Extracting complex visual features is time consuming, with VF running on average at 449.9 millisecond per frame, which is insufficient for performing in real-time. The light-weight pooled bounding box features in RM obtain a processing rate 450 times faster than that achieved in VF, requiring only 1.4 milliseconds per frame.

Considering these values, the required processing time for RM is lower than state-of-the-art trackers such as KCF [11] or real-time RGB-D trackers such as [6], which run respectively on average at 6 and 25 millisecond per frame. Thus, the proposed method can readily fit into real-time monitoring systems. Indeed, the proposed method has been successfully tested in the real-time multi-camera video platform of the SPHERE sensor network system [25].

4 Conclusion

This paper presented a real-time vision system for a contact-free estimation of calorific expenditure estimation in daily living scenarios. The proposed method used pooled temporal pyramids of bounding box features, and subsequently built a recurrent sliding window approach upon it. We demonstrated the effectiveness and efficiency of the proposed method via detailed experiments on accuracy and runtime performance in a comparative study.

The proposed method shows its ability to outperform the widely used METs estimation approach in estimating calorie expenditure, and to provide results in the same region of accuracy of an approach using complex visual features in estimating activity intensity levels, at a fraction of the computational cost. Future work will include investigating fusion approaches for improving prediction results based on both visual and inertial sensors.

Acknowledgements

This work was performed in the SPHERE IRC project funded by the UK Engineering and Physical Sciences Research Council (EPSRC), Grant EP/K031910/1.

References

[1] COSMED K4b2. `http://www.cosmed.com/`.

[2] J.K. Aggarwal and M.S. Ryoo. Human activity analysis: A review. ACM Computing Surveys, 43(3):16, 2011.

[3] J.K. Aggarwal and L. Xia. Human activity recognition from 3D data: A review. Pattern Recognition Letters, 48:70–80, 2014.

[4] B.E. Ainsworth et al. Compendium of physical activities: an update of activity codes and MET intensities. Medicine and Science in Sports and Exercise, 32(9):498–504, 2000.

[5] M. Altini, J. Penders, R. Vullers, and O. Amft. Estimating energy expenditure using body-worn accelerometers: a comparison of methods, sensors number and positioning. IEEE Journal of Biomedical and Health Informatics, 19(1):219–226, 2015.

[6] M. Camplani, S. Hannuna, M. Mirmehdi, D. Damen, A. Paiement, L. Tao, and T. Burghardt. Real-time RGB-D tracking with depth scaling kernelised correlation filters and occlusion handling. Proceedings of the British Machine Vision Conference, pages 145.1–145.11, 2015.

[7] C. Chang and C. Lin. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2(3):27, 2011.

[8] C. Chen, R. Jafari, and N. Kehtarnavaz. Improving human action recognition using fusion of depth camera and inertial sensors. IEEE Transactions on Human-Machine Systems, 45(1):51–61, 2015.

		Prediction (Calories)				Accuracy %				Correlation			
sequence	GT	RM	DM	MET	VF	RM	DM	MET	VF	RM	DM	MET	VF
1	59	75	78	76	*71*	**73.1**	68.4	71.3	*80.2*	0.46	**0.70**	0.66	*0.83*
2	89	87	86	78	*80*	**97.7**	96.5	88.2	*90.3*	**0.74**	0.69	0.57	*0.85*
3	74	74	93	69	*81*	74.9	81.5	**92.7**	*90.1*	**0.90**	0.87	0.63	*0.84*
4	79	52	51	43	*48*	**66.1**	65.2	55.0	*60.4*	**0.84**	0.82	0.78	*0.87*
5	37	46	46	28	*39*	75.1	73.9	**77.6**	*98.6*	**0.87**	0.84	0.77	*0.90*
6	89	86	84	107	*86*	**96.7**	95.2	79.9	*94.3*	**0.83**	0.74	0.63	*0.82*
7	101	97	96	114	*96*	**96.1**	95.1	87.6	*95.3*	0.58	0.39	**0.61**	*0.61*
8	39	40	42	35	*42*	**96.8**	92.8	84.4	*91.9*	**0.72**	0.67	0.57	*0.93*
9	82	79	80	94	*76*	96.7	**98.8**	85.3	*92.8*	**0.81**	0.73	0.71	*0.86*
10	49	79	80	*68*	76	38.9	37.7	**45.2**	*61.5*	**0.58**	0.56	0.42	*0.54*
11	28	44	45	38	*38*	43.5	40.0	66.8	*65.5*	**0.56**	0.45	0.56	*0.64*
12	98	93	91	79	*88*	**95.3**	92.5	80.8	*90.3*	**0.84**	0.73	0.66	*0.56*
13	56	84	82	77	*66*	49.4	52.6	**62.7**	*82.3*	**0.80**	0.75	0.62	*0.78*
14	141	87	87	74	*84*	**61.8**	61.8	52.8	*57.4*	**0.81**	0.71	0.60	*0.86*
15	40	49	48	30	*41*	77.6	**78.7**	74.5	*98.9*	0.84	0.86	**0.94**	*0.94*
16	29	32	33	38	*31*	**89.7**	83.9	69.1	*97.3*	0.71	0.74	**0.81**	*0.88*
17	81	81	82	100	*85*	**99.3**	97.9	76.0	*94.2*	0.38	0.30	**0.70**	*0.74*
18	65	82	83	94	*86*	**73.7**	71.7	54.7	*69.3*	0.25	0.14	**0.48**	*0.83*
19	92	86	85	101	*89*	**93.2**	91.8	90.6	*94.2*	**0.88**	0.75	0.72	*0.75*
20	63	83	84	86	*83*	**67.5**	66.0	64.4	*66.9*	0.53	**0.56**	0.41	*0.81*
Average	-	-	-	-	-	**78.2**	77.1	73.7	*82.9*	0.60	0.56	**0.64**	*0.79*

Table 6. **Estimating Calorific Expenditure.** The table shows ground truth and predicted calorie values in total per sequence and its accuracy and correlation. The best results for each sequence are in bold. (Note that the total calorie value for sequence 5, 8, 11, 15 and 16 are relatively low due to shorter sequences caused by camera errors.)

[9] T.G. Dietterich. Machine learning for sequential data: A review. Structural, Syntactic, and Statistical Pattern Recognition, pages 15–30, 2002.

[10] A. Edgcomb and F. Vahid. Estimating daily energy expenditure from video for assistive monitoring. International Conference on Healthcare Informatics, pages 184–191, 2013.

[11] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista. High-speed tracking with kernelized correlation filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(3):583–596, 2015.

[12] W.D. McArdle, F.I. Katch, and V.L. Katch. Exercise physiology: energy, nutrition, and human performance. Medicine Science in Sports Exercise, 23(12):1403, 1991.

[13] OpenNI organization. OpenNI User Guide, November 2010.

[14] O. Oreifej and Z. Liu. HON4D: Histogram of oriented 4D normals for activity recognition from depth sequences. Computer Vision and Pattern Recognition, pages 716–723, 2013.

[15] F. Perronnin, J. Sánchez, and T. Mensink. Improving the fisher kernel for large-scale image classification. European Conference on Computer Vision, pages 143–156, 2010.

[16] L. Lo Presti and M. La Cascia. 3D skeleton-based human action classification: A survey. Pattern Recognition, 2016.

[17] E. Ravussin, S. Lillioja, T.E. Anderson, L. Christin, and C. Bogardus. Determinants of 24-hour energy expenditure in man. methods and results using a respiratory chamber. Journal of Clinical Investigation, 78(6):1568, 1986.

[18] M.S. Ryoo, B. Rothrock, and L. Matthies. Pooled motion features for first-person videos. Computer Vision and Pattern Recognition, pages 896–904, 2015.

[19] G. Samitz, M. Egger, and M. Zwahlen. Domains of physical activity and all-cause mortality: systematic review and dose–response meta-analysis of cohort studies. International Journal of Epidemiology, 40(5):1382–1400, 2011.

[20] D. Sun, E.B. Sudderth, and H. Pfister. Layered RGBD scene flow estimation. 2015 IEEE Conference on Computer Vision and Pattern Recognition, pages 548–556, 2015.

[21] L. Tao, T. Burghardt, D. Damen, M. Mirmehdi, A. Copper, S. Hannuna, M. Camplani, A. Paiement, and I. Craddock. Calorie counter: RGB-Depth visual estimation of energy expenditure at home. Workshop on Assistive Vision in 13th Asian Conference on Computer Vision, 2016.

[22] L. Tao, T. Burghardt, S. Hannuna, M. Camplani, A. Paiement, D. Damen, M. Mirmehdi, and I. Craddock. A comparative home activity monitoring study using visual and inertial sensors. IEEE International Conference on E-Health Networking, Application and Services, 2015.

[23] L. Tao, A. Paiement, D. Damen, M. Mirmehdi, S. Hannuna, M. Camplani, T. Burghardt, and I. Craddock. A comparative study of pose representation and dynamics modelling for online motion quality assessment. Computer Vision and Image Understanding, 2016.

[24] P-F. Tsou and C-C. Wu. Estimation of calories consumption for aerobics using Kinect based skeleton tracking. International Conference on Systems, Man, and Cybernetics, pages 1221–1226, 2015.

[25] N. Zhu, T. Diethe, M. Camplani, L. Tao, A. Burrows, N. Twomey, D. Kaleshi, M. Mirmehdi, P. Flach, and I. Craddock. Bridging ehealth and the internet of things: The SPHERE project. IEEE Intelligent Systems, 2015.